ing the dataset specification; training and continuing support of those who collect data; and the development and testing of in house database and reporting systems.

To date, 108 units have been recruited—that is, 46% of the units in England and Wales.[5] With a remit to be self financing within two years, the audit is run on a cost recovery basis and makes a small annual charge to each participating unit. It should be noted that the extraction, collection, and entry of accurate data require resources for staff, software, and hardware. The total cost, however, is barely 1% of the annual budget for an average intensive care unit.

The centre was established as a charitable company separate from the society so that it could provide independent and objective audit and research into intensive care. As well as providing feedback to units for audit purposes, the resultant database will be a valuable resource for research. Through the centre's work with the Cochrane Collaboration, identifying and systematically reviewing trials relevant to intensive care, we hope to contribute to the scientific evidence on which future practice in intensive care should be based.

Without a role model to follow, we have learnt several lessons along the way—for example, how to deal with confidentiality and the ownership of data. If the orthopaedic surgeons take up Sochart and colleagues' call we would be happy to share our experiences.

KATHY ROWAN
Director

Intensive Care National Audit and Research Centre,
London WC1H 9HR

1 Sochart DH, Long AJ, Porter ML. Joint responsibility: the need for a national arthroplasty register. *BMJ* 1996;313:66-7. (13 July.)
2 Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's APACHE II study in Britain and Ireland. I. Variations in case mix of adult admissions to general intensive care units and impact on outcome. *BMJ* 1993;307:972-7.
3 Intensive Care National Audit and Research Centre. *ICNARC: an introduction.* London: ICNARC, 1994.
4 Intensive Care National Audit and Research Centre. *Case mix programme dataset specification (ICMPDS).* London: ICNARC, 1995.
5 Metcalfe A, McPherson K. *Study of provision of intensive care in England, 1993. Revised report.* London: Department of Health, 1995.

Current evaluations of information technology in health care are often inadequate

EDITOR,—Two recent articles debated the value of information technology to the NHS and called for further evidence from evaluation studies to inform future investment in information technology.[1 2] These articles draw attention to both the scarcity of evaluation studies of information technology in health care and the lack of scientific rigour of such studies. They reflect current opinion, which strongly advocates the use of economic analyses and randomised controlled trials.

While we agree with the need to provide evidence of the benefits of information technology in health care, we believe that current thinking and practice concerning evaluation are short sighted and detrimental to long term progress. By concentrating on economic analyses of information technology (the results of which have not been encouraging), current evaluation methods overlook potential benefits, such as improved quality of care through better access to more reliable information. By insisting on evidence from randomised controlled trials we waste precious resources on evaluation work that is methodologically flawed and impractical and at best provides results that are difficult or impossible to generalise. An additional problem is that of timing. Many of the benefits of information technology cannot be expected to become evident until several years or more after its introduction, yet we call for evidence of immediate benefits.

In the current climate, in which information technology is viewed suspiciously by many people, it may be argued that evaluation is implicitly used as a barrier to progress. We have inadvertently created a catch 22 situation whereby we cannot move forward with information technology in health care because of the lack of evaluation, yet our failure to deploy systems in routine clinical practice and allow them to mature means that we have nothing of any substance to evaluate.

Information technology is an abstract phenomenon that can be moulded into many forms, often rapidly. It does not operate in a vacuum but is embedded within a complex social and organisational context. The priorities, therefore, should be to develop richer understanding of the effects of its benefits in health care and to develop new evaluation methods that help us to understand the process of implementing it.

Negative results from current inadequate evaluations should not be used to impede general research and development. There is a pressing need to build foundations of fundamental research and development in information technology in health care.

HEATHER A HEATHFIELD
Senior lecturer

Department of Computing,
Manchester Metropolitan University,
Manchester M1 5GD

IAIN E BUCHAN
Research fellow

Medical Informatics Unit,
Institute of Public Health,
University of Cambridge,
Cambridge CB2 2SR

1 Lock C. What value do computers provide to NHS hospitals? *BMJ* 1996;312:1407-10. (1 June.)
2 Donaldson LJ. From black bag to black box: will computers improve the NHS? *BMJ* 1996;312:1371-2. (1 June.)

Placebo mania

Placebos are essential when extent and variability of placebo response are unknown

EDITOR,—We need placebo controlled trials[1] because responses to placebo are remarkably variable. Figure 1 shows the results of randomised double blind trials of anticonvulsants,[2] antidepressants,[3] and topical capsaicin[4] in painful diabetic neuropathy. The proportion of patients in whom the active drug relieved at least 50% of the pain is plotted against the proportion in whom placebo relieved at least 50% of the pain. The active treatments achieved this degree of pain relief in 50-90% of the patients, while placebo did so in 0-80%.

This variation has no simple cause related to the design of the study or to measurement, nor is such variation rare. Trials in other chronic pain conditions showed a placebo response varying from 7% to 86%[2] and from 4% to 47%.[3] In trials of antiemetics 0-50% of patients given a placebo vomited soon after general surgery,[5] and after correction of squint the range of the placebo response was 18% to 88%.[6] In trials of surfactant in the respiratory distress syndrome the range of the placebo response was 24% to 69%.[7]

B Jones and colleagues emphasise the need to know that "both treatments were indeed effective" in a trial of two active drugs.[8] Only if we know the extent of the placebo response and that it does not vary can this criterion be fulfilled. Performing studies of drug A versus drug B when the extent and variability of the placebo response are unknown can be misleading. As Jones and colleagues point out, the results could mean that both A and B were effective or that neither A nor B was effective. The only current defence is to have a placebo group. Designs of trials of analgesics 30 years ago understood this and included both standard active and placebo groups.[9] Only in trials in which sensitivity is proved (that is, standard treatment beats placebo) can correct conclusions of equivalence be made.

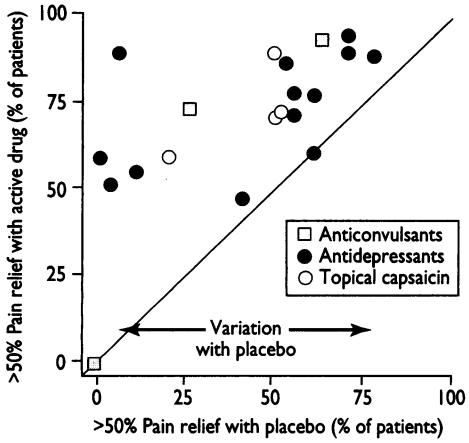

Fig 1—*L'Abbé plot of proportion of patients in whom active drug relieved at least 50% of pain against proportion in whom placebo relieved at least 50% of pain in randomised controlled trials of various drugs in diabetic neuropathy*

If we already know the answer then it is unethical to include a placebo group: the condition of equipoise is not met. If we do not know the answer, and in particular if we do not know the extent and variability of the placebo response, then it is unethical to enrol patients into trials whose design precludes a sensible answer. In these circumstances a placebo group is essential and ethically acceptable.

HENRY MCQUAY
Clinical reader in pain relief
ANDREW MOORE
Consultant biochemist

Pain Research and Nuffield Department of Anaesthetics,
University of Oxford,
Oxford Radcliffe Hospital,
Churchill Headington,
Oxford OX3 7LJ

1 Rothman KJ. Placebo mania. *BMJ* 1996;313:3-4. (6 July.)
2 McQuay H, Carroll D, Jadad AR, Wiffen P, Moore A. Anticonvulsant drugs for management of pain: a systematic review. *BMJ* 1995;311:1047-52.
3 McQuay H, Nye BA, Carroll D, Wiffen PJ, Tramer M, Moore RA. A systematic review of antidepressants in neuropathic pain. *Pain* (in press).
4 Zhang WY, Li Wan Po A. The effectiveness of topically applied capsaicin. A meta-analysis. *Eur J Clin Pharmacol* 1994;46:517-22.
5 Tramer M, Moore RA, McQuay H. Omitting nitrous oxide in general anaesthesia. Meta-analysis of intraoperative awareness and postoperative emesis in randomised controlled trials. *Br J Anaesth* 1996;76:186-93.
6 Tramer M, Moore A, McQuay H. Prevention of vomiting after paediatric strabismus surgery: a systematic review using the numbers-needed-to-treat method. *Br J Anaesth* 1995;75:556-61.
7 Soll JC, McQueen MC. Respiratory distress syndrome. In: Sinclair JC, Bracken ME, eds. *Effective care of the newborn infant.* Oxford: Oxford University Press, 1992:333.
8 Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36-9. (6 July.)
9 Houde R, Wallenstein S, Beaver W. Clinical measurement of pain. In: De Stevens G, ed. *Analgetics.* New York: Academic Press, 1965:75-122.

Placebo controlled trials are needed to provide data on effectiveness of active treatment

EDITOR,—Kenneth J Rothman questions whether there is any point in using a placebo group if blind assessment can be achieved in a comparative trial of two active treatments.[1] He minimises both the practical advantages of placebo controlled trials and their advantages for hypothesis testing. The limitations of statistical

hypothesis testing mean that the probability of an inferential error is known if one is rejecting the null hypothesis but not if one is accepting it.[2] The probability of a type II error in equivalence trials is unknown and can be high with small sample sizes. The conclusion that a difference exists is therefore made on a clearer basis in placebo controlled trials. Equivalence trials are rarely large enough to give the same degree of certainty, and the practical difficulties of obtaining larger sample sizes are more inhibiting than Rothman suggests.

The effect of active treatment is greater in equivalence trials than in placebo controlled trials. For example, in antidepressant trials, placebo controlled trials show a lower efficacy of active drugs than do studies that use no control or an active drug.[3] The context of the trial is important, and the placebo component of the response to treatment seems to be greater in equivalence trials than in placebo controlled trials. Evaluating a new drug only in equivalence trials can lead to an inflated estimate of its effect size.

Blinding may not be as critical in equivalence trials as in placebo controlled trials. Unblinding does occur in clinical trials, leading to the potential for bias.[4] Subjects who have never received any similar treatment are likely to have fewer cues to help them determine whether they are receiving the new or standard treatment in an equivalence trial, but side effects or other cues may help them determine whether they are receiving an active or a placebo drug. Such factors are likely to increase the chance of finding that two active treatments are equivalent, as the degree of unblinding correlates with measured efficacy in a clinical trial.[5] Restricting the evaluation of a new drug to equivalence studies does not give the opportunity to estimate the bias introduced through unblinding.

Too much of the effectiveness of clinical practice is due to the placebo effect for placebo controlled trials to be abandoned. Placebos are required, if only to provide more data about the effectiveness of active treatment. If blind assessments could be achieved in clinical trials and scientific rigour did lead to the accumulation of medical knowledge then more confidence could be placed in equivalence trials.

D B DOUBLE
Consultant psychiatrist

Southeast Sheffield Mental Health Services,
Sheffield S12 4QN

1 Rothman KJ. Placebo mania. *BMJ* 1996;**313**:3-4. (6 July.)
2 Rozeboom AW. The fallacy of the null-hypothesis significance test. *Psychol Bull* 1960;**57**:416-28.
3 Wechsler H, Grosser GH, Greenblatt M. Research evaluating antidepressant medications on hospitalized mental patients: a survey of published reports during a five-year period. *J Nerv Ment Dis* 1965;**141**:231-9.
4 Fisher S, Greenberg RP. How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis* 1993;**181**:345-50.
5 Double DB. Unblinding in trials of the withdrawal of anticholinergic agents in patients maintained on neuroleptics. *J Nerv Ment Dis* 1995;**183**:599-602.

High quality placebos should be used

EDITOR,—In his editorial Kenneth J Rothman states that "as medical knowledge accumulates, the number of placebo trials should fall" and that "placebo control should no longer be part of the gold standard."[1] I disagree with the latter statement. The main purposes of the placebo are to protect evaluations of outcome from (i) the effects of experimenter bias, by enabling blind assessment; and (ii) the influence of patients' expectations, by controlling for non-specific aspects of health care. Such control is particularly important with treatments that depend on subjective assessment of outcome.

While active drugs given in tablet form under double blind conditions may be indistinguishable from each other, there are many other medical procedures that are very different in appearance and mode of administration, for which it would prove difficult to maintain blindness in a comparative trial. Also, the use of a standard treatment in comparative trials would not necessarily control for the same set of non-specific factors as a placebo, which matches the treatment in every way except for the main component believed to be responsible for therapeutic effects—a component control placebo.[2] Indeed, many active treatments produce unique side effects, which may act as a cue to the condition to which participants have been assigned in a clinical trial, thereby leading to unblinding.[3] Evidence suggests that the inclusion of an active as opposed to an inert placebo in such trials would serve to strengthen the double blind paradigm.[4]

Neither should it be assumed that a standard treatment is necessarily more effective than an appropriate placebo rated as being equally as credible as active treatment. A number of orthodox treatments that are currently in use are of unknown efficacy and would not gain ethical clearance if it were sought today. Furthermore, the use of old treatments in comparative trials would serve little in delineating the mechanism by which a new treatment, if it proves effective, achieves its effects.

Therefore, in the hierarchy of evidence of effectiveness, a randomised controlled trial that uses a component control, active, or credible placebo should feature higher than a comparison with an active drug or medical procedure of unknown effectiveness. The gold standard could be refined by the use of these high quality placebos in place of the standardised "sugar pill" placebo.

AMANDA GEORGIOU
Lecturer in public health

Division of Public Health,
Nuffield Institute for Health,
University of Leeds,
Leeds LS2 9PL

1 Rothman KJ. Placebo mania. *BMJ* 1996;**313**:3-4. (6 July.)
2 Horvath P. Placebos and common factors in two decades of psychotherapy research. *Psychol Bull* 1988;**104**:214-25.
3 Greenberg RS, Bornstein RF, Zborowski MJ, Fisher S, Greenberg MD. A meta-analysis of fluoxetine outcome in the treatment of depression. *J Nerv Ment Dis* 1994;**182**:547-51.
4 Fisher S, Greenberg RP. How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis* 1993;**181**:345-50.

Use of anabolic steroids has been reported by 9% of men attending gymnasiums

EDITOR,—F H Lloyd and colleagues report several cases of subfertility in male body builders who misused anabolic steroids.[1] I would draw the authors' attention to a report commissioned by the Departments of Health for England, Scotland, and Wales, entitled *Anabolic Steroid Use in Great Britain: An Exploratory Investigation.*[2]

The research for the report investigated use of anabolic steroids in 21 gymnasiums in England, Scotland, and Wales and found that 119 (9.1%) of the 1310 male respondents to the questionnaire and eight (2.3%) of the 349 female respondents had taken anabolic steroids. The youngest user was aged 16. The prevalence of use of anabolic steroids in the gymnasiums ranged from zero (in three gymnasiums) to 46% (28 of 61 respondents). The response rate to the questionnaire was 59% (1677/2834). These findings were confirmed by a similar study by Lenehan *et al*, who investigated the prevalence of use of anabolic steroids in gymnasiums in Merseyside.[3]

Lloyd and colleagues express concern about male subfertility and raise several important points, including users' lack of knowledge of possible side effects, the high doses taken, and polypharmacy. These concerns were confirmed in in depth interviews with 110 users of anabolic steroids (97 men and 13 women).[2] Fifty four (56%) of the men reported having experienced testicular atrophy, and eight (62%) of the women reported menstrual irregularities, which suggests that subfertility is also a problem among women. Only 36 of the users had told their general practitioner that they used anabolic steroids. Polypharmacy was common: 55 of the interviewees took three or more different drugs during their "on cycle" of anabolic steroids, and 88 injected them. Use of several other drugs, including human chorionic gonadotrophin, oestrogen antagonists, growth hormone, thyroid hormones, and nalbuphine hydrochloride, was also reported. The findings also suggested that use of other substances is not uncommon: in the previous six months 20 of the interviewees had taken amphetamines, 26 cannabis, and four cocaine. Forty eight reported that dealers were their main sources of anabolic steroids and 28 that friends were; dealers and friends were also the main sources of information about anabolic steroids, along with the "underground" handbooks on steroids.

PIRKKO KORKIA
Senior lecturer in sport and exercise science

University of Luton,
Luton LU1 3JU

1 Lloyd FH, Powell P, Murdoch AP. Anabolic steroid abuse by body builders and male subfertility. *BMJ* 1996;**313**:100-1. (13 July.)
2 Korkia P, Stimson GV. *Anabolic steroid use in Great Britain: an exploratory investigation. A report to the Department of Health, the Welsh Office and the Chief Scientist Office, Scottish Home and Health Department.* London: HMSO, 1993.
3 Lenehan P, Bellis M, McVeigh J. *Anabolic steroid use in the north west of England. A summary.* Liverpool: Drugs and Sport Information Service, 1996.

Usefulness of urine dipstick tests

False negative results may occur in the absence of antibiotics, ketones, and glucose

EDITOR,—Several studies,[1 2] including that by J H Beer and colleagues,[3] have highlighted the fact that there is still debate over the diagnostic accuracy of rapid dipstick tests. Beer and colleagues point out that the presence of certain antibiotics in urine samples may cause false positive results. Other workers have pointed out that false negative results may be due to antibiotics, glucose, or oxalic acid. Bonnardeaux *et al* concluded that results of dipstick tests were an unreliable predictor of results of microscopy and that false negative results can result from the presence of ketones or glucose.[2] Hurlbut and Littenburg concluded that "a negative urine dipstick test cannot exclude the diagnosis of urinary tract infection."[1]

In a recent study of 1705 clinical urine specimens performed in our laboratory we compared automated (Clintec 200+, Ames) and manual reading of dipsticks (Multistix 8SG in both cases) with quantitative leucocyte determinations made with microchamber slides and a standardised urine analysis system (Kova disposable slides, Hycor Biomedical). We also tested each urine specimen with a variable orifice particle analyser (Questor, Micro-Med). To distinguish positive from negative samples we used a cut off point of 70 white cells/ml for all tests except analysis with the variable orifice particle analyser (for which we used a cut off of 100 white cells/ml). To overcome variability among the test methods a specimen had to give a positive result